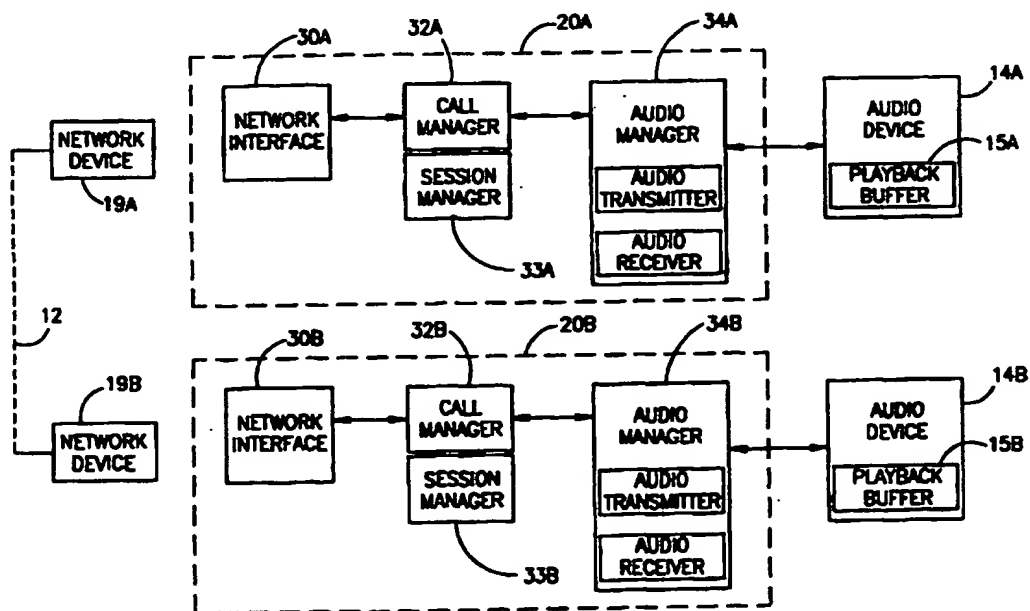




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : <b>H04J 3/02, 3/12, 3/16, 3/24, H04Q 11/04</b>		<b>A1</b>	(11) International Publication Number: <b>WO 96/15598</b>
			(43) International Publication Date: 23 May 1996 (23.05.96)
(21) International Application Number: PCT/US95/14123		(81) Designated States: AL, AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UG, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, LS, MW, SD, SZ, UG).	
(22) International Filing Date: 9 November 1995 (09.11.95)		<p><b>Published</b> <i>With international search report.</i></p>	
(30) Priority Data: 337,616 10 November 1994 (10.11.94) US			
(71) Applicant: VOCALTEC, INC. [US/US]; 157 Veterans Drive, Northvale, NJ 07647 (US).			
(72) Inventors: COHEN, Alon; 16 Har-Zion Street, 75238 Rishon LeZion (IL). HARAMATY, Lior; 20 McDonald Street, 52514 Ramat Gan (IL).			
(74) Agents: GUNNISON, Forrest, E. et al.; Skjerven, Morrill, MacPherson, Franklin & Friel, Suite 700, 25 Metro Drive, San Jose, CA 95110 (US).			

(54) Title: AN AUDIO TRANSCIVER



## (57) Abstract

The present invention is an audio transceiver (20) which on the receiving side (20B) controls the amount of audio data in the buffer (15B) of a PC audio device (14B), such that the audio device (14B) always has something to play. On the transmission side (20A), the audio transmitter provides at least sequence numbers to the audio packets to be sent. The audio receiver is concerned only with the state of the buffer of the audio device. Therefore, the audio transmitter does not have to be synchronized with the audio receiver.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgyzstan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LU	Luxembourg	TD	Chad
CS	Czechoslovakia	LV	Latvia	TG	Togo
CZ	Czech Republic	MC	Monaco	TJ	Tajikistan
DE	Germany	MD	Republic of Moldova	TT	Trinidad and Tobago
DK	Denmark	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	US	United States of America
FI	Finland	MN	Mongolia	UZ	Uzbekistan
FR	France			VN	Viet Nam
GA	Gabon				

## AN AUDIO TRANSCEIVER

### FIELD OF THE INVENTION

The present invention relates to apparatus for providing real-time or near real-time communication of audio signals via a data network.

### BACKGROUND OF THE INVENTION

Data networks transfer data, typically in the form of packets which usually have a fixed number of bytes of data, from one workstation to another. There are many types of network protocols by which networks setup communication paths. Ethernet and Token Ring are examples of low level network structures for packet data networks.

Regardless of the type of structure used, no network instantaneously provides packets from a source workstation to a destination one. There is a transmission delay which typically varies depending on the load on the network (i.e. how many workstations are trying to send at once) and/or on the configuration of the network (i.e. which path the packet takes) and type of protocol used.

If two, sequential packets take two different paths through the network, it is possible that they will arrive at the destination workstation after different amounts of time traveling through the network. They also might possibly arrive in the wrong order. Since most data transmitted over a network is transmitted for storage purposes, the delays and mixed up order are not critical, although it is always desirable to reduce them to a minimum.

Audio devices, which convert analog audio signals to digital ones, are known. These devices sample the analog audio signal, at some sampling frequency, to produce a digital datastream and then compress the datastream to reduce the storage or bandwidth requirements for storing or transmitting the datastream. The datastream can then be divided into packets and transmitted along a network, to be reassembled and played by the destination workstation. The playing involves converting the packets into the datastream which is then converted back into an analog signal. As is known in the art, digital to analog conversion also involves a converting frequency which is typically the same frequency as the sampling frequency of the audio device.

If the audio signal is to be stored by the destination workstation, then the delays and changed sequence are not critical. When the audio signal is retrieved from storage and played, it will be played smoothly since all of its packets are present in the storage medium.

5           However, if a real-time conversation is desired, an "audio packet" should be played by the audio device as soon as it arrives. This is difficult when working over a network for exactly the reasons described hereinabove; the packet order is not necessarily maintained during transmission and there is a network delay which is not of a fixed value. Furthermore, even if the delays are overcome, if the audio device of the source workstation has a sampling  
10 frequency which is different (faster or slower) than the converting frequency of the audio device of the destination workstation, the two cards will not be synchronized. If the source workstation samples at a higher frequency, the destination workstation will not be able to play the packets fast enough. Conversely, if the source workstation samples at a lower frequency, the destination workstation will not have enough packets to play.

15           The following two articles discuss the issues involved in providing audio communication over a packet data network:

          Clifford J. Weinstein and James W. Forgie, "Experience with Speech Communication in Packet Networks", IEEE Journal on Selected Areas in Communications, Vol. SAC-1, No. 6, December 1983, pp. 963 - 980; and

20           Warren A. Montgomery, "Techniques for Packet Voice Synchronization", IEEE Journal on Selected Areas in Communications, Vol. SAC-1, No. 6, December 1983, pp. 1022 - 1028.

          The first article discusses network protocols for transmitting speech. The second article discusses a packet voice receiver unit which chooses a target playout time for  
25 each packet. The playout time is a fixed interval after its production by the source workstation. The packet is played only if it arrives before its target playout time. The second article also discusses a number of methods for determining the delay encountered by a packet due to the network. Since the second article assumes that the two audio devices are almost synchronized (i.e. their frequencies are very close) and that speechbursts are short,  
30 it increases the target playout time to compensate for the lack of synchronization.

          The second article also discusses adaptively changing the target playout time, typically during silent periods. It can also change the target playout time during playout, although the article mentions that changing the playout time during playout requires

maintaining the pitch of the speech. Finally, the second article discusses the impact of synchronization techniques on network design.

Programs for enabling audio communication over networks of similar types of workstations are known. For example, the programs NetFone and Vtalk are designed to send voice signals over a data network; however, these programs work only between workstations manufactured by Sun Microsystems, Inc. of USA.

A voice communication system over a network running the Ethernet protocol is commercially available from Genisys Comm Inc. of Rome, New York, USA. This system works with personal computers (PCs).

## SUMMARY OF THE PRESENT INVENTION

It is an object of the present invention to provide an audio transceiver between a personal computer (PC) and a packet data network.

5       The present invention is an audio transceiver (having an audio receiver and a transmitter) which, on the receiving side, adaptively controls the amount of audio data in the buffer of a PC audio device, such that the audio device always has something to play. On the transmission side, the audio transmitter provides at least sequence numbers to the audio packets to be sent. The audio receiver receives the audio packets, processes them and plays  
10       them as soon as possible thereafter.

      Since the present invention does not measure the amount of time it took for the audio packets to come, the audio receiver and transmitter can be placed at the ends of any size network (one with a short delay or one with a long delay).

15       In addition, the audio receiver is concerned only with the state of the buffer of the audio device. Therefore, the audio transmitter does not have to be synchronized with the audio receiver. Their clocks can be slightly or significantly different; the audio receiver can handle both situations.

20       Specifically, in accordance with a preferred embodiment of the present invention, the audio transceiver includes, apparatus for sequence-stamping outgoing audio packets received from the audio device and, on input, apparatus for receiving a stream of the sequence-stamped audio packets from the packet data network, fullness setting apparatus and fullness adjusting apparatus. The fullness setting apparatus transfers a silence buffer to a playback buffer of the audio device whenever the playback buffer is empty. The fullness adjusting apparatus adaptively controls the fullness of the playback buffer to generally match  
25       the playout rate of the audio device with the rate at which the audio packets are received.

      In addition, in accordance with a preferred embodiment of the present invention, the transceiver includes apparatus for sequence- and destination-stamping all of the audio packets and apparatus for transmitting the audio packets via the network.

30       Moreover, in accordance with a preferred embodiment of the present invention, the transceiver includes sound detection apparatus which receives audio packets from the first audio device, which determines when the audio packets begin to contain sound and which sends the audio packets from the beginning of the sound.

      Still further, in accordance with a preferred embodiment of the present invention, the packet data network is a private or, alternatively, a public network.

Additionally, in accordance with a preferred embodiment of the present invention, the fullness setting apparatus includes apparatus for increasing an adjustable fullness level. The adjusting apparatus includes apparatus for decreasing the adjustable fullness level and apparatus for processing audio data within the audio packets in order to fill  
5 the playback buffer to the current value of the fullness level.

Moreover, in accordance with a preferred embodiment of the present invention, the apparatus for processing includes apparatus for determining the amount of data in the playback buffer during a predetermined window of time. The apparatus for processing typically includes apparatus for adding and removing portions of the audio data as a function  
10 whether or not the current amount of data is less or more than the current value of the fullness level.

Further, in accordance with a preferred embodiment of the present invention, the apparatus for adding and removing includes apparatus for maintaining the size of the portions of audio data until the current amount of data reaches the current value of the  
15 fullness level.

Finally, the present invention includes a method for processing audio data which includes the actions performed by the elements described hereinabove.

**BRIEF DESCRIPTION OF THE DRAWINGS**

The present invention will be understood and appreciated more fully from the following detailed description taken in conjunction with the drawings in which:

5           Fig. 1 is a schematic illustration of a plurality of workstations connected together via a network, wherein each workstation has an audio transceiver constructed and operative in accordance with a preferred embodiment of the present invention;

          Fig. 2 is a block diagram illustration of the elements of the audio transceiver shown in Fig. 1;

10           Fig. 3 is a schematic illustration of the transmitter portion of the transceiver of Fig. 2 in conjunction with the audio device;

          Fig. 4 is a block diagram illustration of receiver elements of the audio transceiver responsible for reassembling the audio packets received from the network such that they can be played in real-time; and

15           Fig. 5 is a flow chart illustration of the method performed by the receiver of Fig. 4.



## DETAILED DESCRIPTION OF A PREFERRED EMBODIMENT

Reference is now made to Fig. 1 which illustrates a network having audio communication via a plurality of audio transceivers, constructed and operative in accordance with a preferred embodiment of the present invention, and to Fig. 2 which illustrates, in general block diagram format, the elements of one audio transceiver of the present invention.

Fig. 1 illustrates a plurality of workstations 10 connected together via a packet data network 12. The data network 12 can be any type of network, such as a local area network (LAN) or a wide area network (WAN), and it can run any desired network protocol, such as SPX/IPX, TCP/IP, etc. Each workstation 10 is formed of a personal computer (PC) having an audio device 14 and a network device 19. The network device 19 connects its workstation 10 to the network 12. The audio device 14 is connected to a speaker 16 and a microphone 18 and is operative to play digitally recorded sound on the speaker 16 and to convert sound from the microphone 18 to a digital signal. Typical audio devices 14 have playback buffers 15.

The audio transceivers 20 of the present invention bridge between the audio devices 14 and the network devices 19 so that two workstations 10 can provide sound to each other in real- or near real-time, thus enabling the users at the two workstations 10 to have a reasonable voice conversation with each other.

As will be described in more detail hereinbelow, the audio transceiver 20 has an audio receiver and an audio transmitter. On the transmission side, the audio transmitter converts the audio datastream to packets and provides at least sequence numbers to the packets. On the receiving side, the audio receiver receives the audio packets and, in accordance with a preferred embodiment of the present invention, adaptively controls the amount of audio data in the playback buffer 15 of the audio device 14 to maintain a desired fullness level.

Fig. 2 illustrates the general structure of two audio transceivers 20, a source transceiver 20a and a destination transceiver 20b. The explanation of the general operation of the audio transceiver 20 will be provided herein in the context of a conversation between transceivers 20a and 20b.

Each transceiver comprises a network interface 30, a call manager 32, an audio manager 34, and a session manager 33, where the elements of the source transceiver 20a are labeled with an 'a' suffix and those of the destination transceiver are labeled with a 'b' suffix. The network interfaces 30 divide the audio datastream into packets and, via the

network device 19, the network interfaces 30 connect to the network 12. The network interfaces 30 know the addresses of the workstations on the network and serve to connect their audio transceivers 20 to the desired destination workstation.

5 Through the source call manager 32a, the operator indicates with whom he wants to talk. The source call manager 32a converts the name of the person to the address of the workstation at which the person works and prepares a "call initiation" message (a data message) to that destination workstation. The source network manager 30a sends the call initiation message. The destination network manager 30b receives the call initiation message and provides it to its call manager 32b which, in turn, indicates to its operator that a call is  
10 being initiated. This indication can be via the display of the destination PC or by making a "call initiation" sound, such as that of a bell, on the destination audio device. Typically, the destination call manager 32b also indicates to the operator who initiated the call.

The operator, if he wishes to talk to the person who initiated the call, makes an appropriate indication to the destination call manager 32b. In response, the destination  
15 call manager 32b sends an "OK to talk" message, through its network interface 30b, to the source audio transceiver 20a. The "OK to talk" message also indicates to the destination network interface 30b that further messages (which will contain audio data) are to be sent to its audio manager 34b.

The source network interface 30a, upon receipt of the "OK to talk" message,  
20 sends it to the source call manager 32a which, in turn, may provide an appropriate indication to its operator. The indication can be any desired type of indication, such as a sound like a telephone being picked up or some phrase, such as "OK to talk" or "Open", which indicates that the call has been successfully initiated. The "OK to talk" message also indicates to the source network interface 30a that any further messages are to be communicated to and from  
25 the audio manager 34a.

The call managers 32a and 32b periodically send call control signals indicating that their audio transceiver is currently active. The managers 32a and 32b monitor the flow of these control signals and also provide "end of conversation" indications. These can come as commands from the respective operators or after a predetermined length of time during  
30 which no control signal was received from the destination transceiver 20b.

The session managers 33 provide overall control to the elements of each audio transceiver 20. In particular, they manage the logical level of the session with the remote party.

The audio managers 34a and 34b process the digital audio data received from their respective network interfaces 30 and from their respective audio devices 14. The audio managers 34 are divided into audio transmitters 35 (detailed in Fig. 3) and audio receivers (detailed in Figs. 4 and 5).

5 During a conversation, the transmitting audio manager 34a receives the audio datastream from its corresponding audio device 14 and processes the datastream to remove any silent parts. The resultant datastream is provided to the network interface 30a which divides the datastream into packets and adds network information, such as source and destination workstation addresses, to each packet. The packets are then sent to the network  
10 12.

The receiving network interface 30b receives the packets and strips them of the network information, producing thereby an audio datastream. The receiving audio manager 34b processes the datastream in order to ensure that the playback buffers, labeled 15a and 15b, of their respective audio devices 14 have enough digital audio data to play,  
15 irrespective of a) the rate at which the packets arrive, b) the sampling rate of the source audio device 14a or c) the time at which the packets were originally produced.

If desired, the audio manager 34a can compress the audio datastream prior to sending it to the network interface 30a to form into packets. The compression (and decompression on the reception side) can be implemented using any suitable audio  
20 compression/decompression technique, such as the Adaptive Delay Pulse Code Modulation (ADPCM) technique described in the CCITT G.721 standard.

Fig. 3, to which reference is now made, illustrate the elements and operation of the audio transmitter 35 of one audio manager 34. The audio transmitter 35 comprises a generally lossless sound detector, formed of a voice operated transmitter (VOX) 40, a buffer  
25 42 and switch means 44. The sound detector removes any silent periods and enables the users to speak without having to indicate when he is finished speaking (i.e. so that the other person can begin speaking).

It is noted that people do not talk continuously but rather talk in bursts, known as "speechbursts". The sound detector determines when the audio datastream includes a  
30 speechburst (as opposed to background noise) and shifts the datastream to account for the processing time of the VOX 40. Thus, the datastream which the VOX 40 processes is also stored in the buffer 42 whose length is generally related to the processing time of the VOX 40. Once the VOX 40 detects a significant sound within the datastream (which typically occurs near but not at the beginning of a speechburst), it indicates to the switch means 44

to output the data stored in the buffer 42. If no sound was detected, the data stored in buffer 42 are overwritten.

In particular, the VOX 40 considers sound to be present as soon as some data within the buffer 42 is above a typically, but not necessarily, user-adjustable, sound threshold level. The entire contents of the buffer 42 (the datapoint above the sound threshold level plus all of the data before it), are output to the network interface 33 for division into packets.

When the buffer has had no datapoints above a silence threshold level, which is typically lower than the sound threshold level, for a few milliseconds (i.e. the speechburst or conversation has ended), the VOX 40 indicates to the switch means 44 that to disconnect the buffer 42 from the network interface 33.

Reference is now made to Fig. 4 which illustrates the elements of the audio receiver 37 of one audio manager 34. Audio receiver 37 comprises a packet handler 50, an initial fullness setting unit 51, a fullness adjuster 52, and switching means (noted by switches 54) for switching between the units 51 and 52. The output of audio receiver 37 is provided to the playback buffer 15 of the audio device 14.

It is noted that, since people speak in speechbursts, once a speechburst has ended, the playback buffer 15 will have nothing left to play. Thus, the fullness setting unit 51 is activated at the beginning of each speechburst.

It is also noted that the playback buffer 15 is a first-in, first-out (FIFO) buffer which, when requested by the audio device 14, provides the audio device with the oldest audio data stored therein. There is a minimum level of fullness, which varies with the type of audio device 14 utilized, below which the playback buffer 15 should not go, except if the speechburst has ended.

The packet handler 50 receives the audio datastream and the sequence number of each packet from the corresponding network interface 30 and notes the sequence number of the packet. It is noted that each packet stores a plurality of "frames" of audio data and that each frame of audio data can be of any length and can include compressed or uncompressed data in it.

The packet handler 50 resamples the audio data to match the converting frequency of its corresponding audio device 14, as described in more detail hereinbelow. Packet handler 50 also compensates for missing packets by utilizing the packets before and after the missing packets and, if necessary, by adding frames of silence. Frames of silence are frames with silence sounds in them.

Since the network routes each packet separately, the packets do not, necessarily, arrive in order or at a regular rate. At the beginning of a speechburst, this "jitter" in the arrival rate can be extremely problematic. Therefore, the fullness setting unit 51 determines a desired fullness level to overcome most of the jitter and provides the playback buffer 15 with a block of silence data to fill the playback buffer 15 to the desired fullness level. The fullness unit 51 indicates to the adjuster 52 what the fullness level is, after which, the switch means 54 switch control to the fullness adjuster 52.

While the audio device 14 is playing the silence block, the fullness adjuster 52 handles the incoming audio data and provides them to the playback buffer 15. The fullness adjuster 52 adds or removes audio data in order to match the playback rate of the audio device 14 with the rate at which the converted audio data is present. In other words, the packet handler 50 generally converts, or scales, the audio datastream to the converting rate of the audio device 14 of the destination workstation and the fullness adjuster 52 performs fine adjustments to the data rate of the incoming datastream to more accurately match the converting rate of the audio device 14. To do so, the fullness adjuster 52 adjusts the desired fullness level.

If the audio device 14 plays all the data in its buffer 15, either before or when the speechburst ends, typically due to increased jitter on the network, the switch means 54 switches control to the fullness setting unit 51 which slightly increases the fullness level to compensate for the increased jitter and provides another silence block of the size of the increased fullness level.

Fullness adjuster 52 processes the audio data to ensure that the playback buffer 15 is as full as necessary but not overly full, since the more data stored in the playback buffer 15, the longer it takes before the operator hears the received data. Fullness adjuster 52 adjusts the rate of the audio data so as to generally match the playback rate of the audio device 14. Thus, if the playback rate is faster than that of the converted audio data, fullness adjuster 52 adds extra samples to the audio data every so many samples. Conversely, if the playback rate is slower than the rate of converted data, fullness adjuster 52 drops every so many audio samples.

It will be appreciated that the packet handler 50 and the fullness adjuster 52 not only compensate for mismatches between the playback rate and the rate at which the network transfers packets, but also compensate for differences between the packet creation rate of the source audio device 14a and that of the playback rate of the destination audio device 14b. Thus, the audio transceiver of the present invention enables communication

between PCs having audio devices by different manufacturers, which typically do not have similar sampling and playback rates. Similarly, the present invention enables a single audio device manufacturer to produce audio devices whose sampling rates range within a large tolerance range.

5 It will further be appreciated that the fullness setting unit 51 and the fullness adjuster 52 operate to maintain the playback buffer 15 full, without any knowledge of when the incoming audio data were originally produced.

Reference is now made to Fig. 5 which illustrates the operation of the receiver 37, for each incoming packet, in flow chart format.

10 When a packet arrives (step 100), its datastream is first resampled (step 102) to match the converting frequency of the destination audio device 14. The resampling procedure can be any resampling procedure which performs anti-alias filtering, interpolation and decimation. The method utilized by the CAT audio device, commercially available from the common assignees of the present invention, is suitable and is operative on PCs. Other  
15 methods of resampling are also known.

Afterwards, the sequence number of the new packet is compared to that of the previously received packet. If there is a gap between the two sequence numbers (step 104), the gap is filled (step 106). One method for filling the gap is as follows: The frames bordering each side of the gap are duplicated and the remaining frames of the missing packet  
20 or packets are filled with silence. Thus, if the two received packets have frames P1, P2, P3 and P4, P5, P6 in order, and one packet is missing, the resultant series will be: P1, P2, P3, P3, silence, P4, P4, P5, P6. Other methods of filling the gap are also possible.

Steps 100 - 106 form the operations of packet handler 50.

25 Whether or not a packet was missing, in step 108, the receiver 37 determines the current amount of data DATA\_AMOUNT stored in the playback buffer. The current amount of data DATA\_AMOUNT indicates the delay between the arrival of an audio sample and the time it is played out by the audio device 15 and is defined as the difference between the amount of data sent for use by the playback buffer 15 and the amount of data which the playback buffer 15 utilized. In equation format:

30

$$\text{DATA\_AMOUNT} = \text{AMOUNT\_SENT} - \text{AMOUNT\_RETURNED} \quad (1)$$

Typically, the amount sent and amount returned are continually calculated over the course of a speechburst. DATA\_AMOUNT is calculated over a moving window of time (of typically 2 seconds) and thus, is an average, rather than an instantaneous value.

5 In step 110, the receiver 37 determines if the current amount of data DATA\_AMOUNT is 0 (i.e. the playback buffer 15 is empty). If it is more than 0, step 114 is performed. Otherwise, step 112 is performed.

10 If, despite the operation of the fullness adjuster 52, the playback buffer 15 was emptied, indicates that the network jitter has gotten worse or that the speechburst has ended. To compensate for the possible increased jitter, the fullness setting unit 51 increases the fullness level by a predefined amount, such as by 10%. At the same time, the fullness setting unit 51, in step 112, sends a silence block of at least the size of the current desired fullness level.

15 If, alternatively, the playback buffer was not empty, the fullness adjuster 52 is operative. In step 114, it determines whether or not the current desired fullness level is too large and in step 116, it adjusts the data to be sent to the playback buffer 15 in order to achieve the desired fullness level.

20 The current desired fullness level is increased (in step 112), whenever the playback buffer 15 approaches empty and is decreased (in step 114) whenever there were no gaps during the last predetermined length of time, such as for 10 seconds. If the minimum current amount of data DATA\_AMOUNT for the last, say 10 seconds, is larger than the minimum allowed for the specific audio device 14, then the desired fullness level is set to the mean value between the minimum allowed amount of data (for the specific audio device 14) and the minimum DATA\_AMOUNT for the last, say, 10 seconds.

25 It will be appreciated that any other function to reduce the desired fullness level which reduces the level without causing a gap to occur, is also suitable.

30 In step 116, the difference between the current amount of data DATA\_AMOUNT and the desired fullness level is determined. The difference can be measured in seconds of data or in numbers of blocks of data. The data to be sent to the playback buffer 15 is then processed to force the difference to be as close to zero as possible.

The processing involves adding or removing audio samples as a function of how large the difference is. A positive difference (i.e. the current amount of data is larger than the desired fullness level), indicates that the input rate is higher than the playback rate.

Therefore, some of the audio samples should be removed. A negative difference requires the addition of audio samples.

For each type of audio device, the receiver 37 has a LookUp Table (LUT) defining the function for adding and removing. The following table is useful for the Soundblaster audio devices:

Difference (in msec)	Duplication amount
500	-1 per 20 = -5%
300	-1 per 50 = -2%
150	-1 per 100 = -1%
-150	+1 per 100 = +1%
-300	+1 per 50 = +2%
-500	+1 per 20 = +5%

where +1 and -1 indicate addition and removal of a frame and "per X" indicates for every X frames.

When there are frames which have been artificially added due to a missing packet (in step 106), then the artificially added frame is selected as the one to be removed. When a frame is to be added, the frame which is added is a copy of the frame which will be next to it.

Although not shown in Fig. 5, the particular duplication amount is maintained until the difference is close to zero. At that point, the duplication amount can be changed. Furthermore, when duplication or removal is occurring, the window size for determining the current amount of data is reduced, for example to 500 msec.

Once the processing has finished, the data of the packet are sent (step 118) to the playback buffer 15 and the process repeated for the next packet.

The following pseudo code details the operation of the duplication/removal mechanism of step 116:

PseudoCode for Step 116:

PDup - Previous Duplication\_Amount



Duplication\_Amount

(Duplication\_Amount<0, means deletion).

(Duplication\_Amount>0, means duplication).

5 Data\_Amount-(averaged over 2 sec.)

Short\_Data\_Amount-(averaged over 500 ms.)

Fullness\_Level

Difference\_in\_Amount

10 Init:

Duplication\_Amount=0;

PDup=0;

When packet arrives:

15 (Reevaluate Duplication\_Amount)

Difference\_in\_Amount = Data\_Amount - Fullness\_Level

Short\_Difference\_in\_Amount = Short\_Difference\_in\_Amount - Fullness\_Level

Duplication\_Amount = find (Difference\_in\_Amount) in table.

20 (Stop Dup./Del Process)

if (PDup! = 0) {

if (PDup < 0 && Short\_Difference\_in\_Amount < 0)

PDup = 0; (Stop Deletion)

if (PDup > 0 && Short\_Difference\_in\_Amount > 0)

25 PDup = 0; (Stop Duplication)

if (|Duplication\_Amount| > |PDup|)

PDup = Duplication\_Amount;

if difference increases then increase

Dup./Del. accordingly)

30 }

if (PDup == 0){

Duplication\_Amount = find(Difference\_in\_Amount) in table.

PDup = Duplication\_Amount

}

It will be appreciated by persons skilled in the art that the present invention is not limited to what has been particularly shown and described hereinabove. Rather the scope of the present invention is defined by the claims which follow:

## CLAIMS

1. An audio transceiver between an audio device of a personal computer and a packet data network, the audio transceiver comprising:
  - 5 a. means for receiving a stream of sequence stamped audio packets from said packet data network;
  - b. fullness setting means for transferring at least a silence frame to a playback buffer of said audio device whenever said playback buffer is empty in order to fill said playback buffer to an adjustable fullness level;  
10 and
  - c. fullness adjusting means for controlling the fullness level of the playback buffer, said fullness adjusting means operating to duplicate or remove at least one frame until a difference between a current fullness level and the desired fullness level is substantially zero.
- 15 2. A transceiver according to claim 1 and also comprising:
  - a. means for sequence- and destination-stamping all of said audio packets;  
and
  - b. means for transmitting said audio packets via said network.
- 20 3. A transceiver according to any of claims 1 - 2 and also comprising voice detection means which stores at least a portion of said received audio packets and sends after said sound threshold level is reached said stored audio packets with audio packets received after said sound threshold level is reached.
4. A transceiver according to any of claims 1 - 3 and wherein said packet data network is a private network.
- 25 5. A transceiver according to any of claims 1 - 3 and wherein said packet data network is a public network.
6. A transceiver according to any of the previous claims and wherein said fullness setting means comprises means for increasing said adjustable fullness level and wherein said fullness adjusting means includes means for decreasing said

adjustable fullness level and means for processing audio data within said audio packets in order to fill said playback buffer to the current value of said adjustable fullness level.

- 5 7. A transceiver according to claim 6 and wherein said means for processing comprises means for determining the amount of data in said playback buffer during a predetermined window of time.
- 10 8. A transceiver according to any of claims 6 - 7 and wherein said means for processing includes means for adding and removing portions of said audio data as a function of whether or not the current amount of data is less or more than said current value of said adjustable fullness level.
- 15 9. A transceiver according to claim 8 and wherein said means for adding and removing comprise means for maintaining the size of said portions of audio data until said current amount of data reaches said current value of said adjustable fullness level.
- 20 10. A transceiver according to the previous claims wherein said means for receiving further comprising means for duplicating at least one frame of two packets bordering a gap, said gap detected whenever the sequence number of said two adjacent packets is not consecutive.
11. A transceiver according to claim 10 wherein said means for receiving further comprises inserting at least one silence frame between said duplicated frames.
12. A transceiver according to any of claims 10 - 11 wherein said fullness adjusting means remove at least one frame of the frames duplicated by said means for receiving due to said gap.
- 25 13. A transceiver according to any of claims 10 - 12 wherein said fullness adjusting means duplicate at least one frame of the frames not duplicated by said means for receiving due to said gap.

14. A transceiver according to any of claims 10 - 12 wherein said adjustable fullness level is determined in accordance with the variations in size of said playback buffer.

5 15. A method for transmitting and receiving audio data between an audio device of a personal computer and a packet data network, the method comprising the steps of:

a. receiving a stream of sequence-stamped audio packets from said packet data network;

10 b. transferring at least a silence frame to a playback buffer of said audio device whenever said playback buffer is empty in order to fill said playback buffer to an adjustable fullness level; and

15 c. controlling the fullness level of the playback buffer, said controlling comprising duplicating or removing at least one frame until a difference between a current fullness level and the desired fullness level is substantially zero.

16. A method according to claim 15 and also comprising the steps of:

a. sequence- and destination-stamping all of said audio packets; and

b. transmitting said audio packets via said network.

20 17. A method according to any of claims 15 - 16 and also comprising the steps of storing at least a portion of audio packets received before a sound threshold level is reached and sending after said sound threshold level is reached said stored audio packets with audio packets received after said sound threshold level is reached.

25 18. A method according to any of claims 10 - 17 and wherein said packet data network is a private network.

19. A method according to any of claims 10 - 17 and wherein said packet data network is a public network.

- 5 20. A method according to any of claims 10 - 19 and wherein said step of transferring includes the step of increasing said adjustable fullness level and wherein the step of controlling includes the steps of decreasing said adjustable fullness level and processing audio data within said audio packets in order to fill said playback buffer to the current value of said adjustable fullness level.
21. A method according to claim 20 and wherein said step of processing includes the step of determining the amount of data in said playback buffer during a predetermined window of time.
- 10 22. A method according to any of claims 20 - 21 and wherein said step of processing includes the step of adding and removing portions of said audio data as a function of whether or not the current amount of data is less or more than said current value of said adjustable fullness level.
- 15 23. A method according to claim 22 and wherein said step of adding and removing includes the step of maintaining the size of said portions of audio data until said current amount of data reaches said current value of said adjustable fullness level.
- 20 24. A method according to any of claims 20 - 23 wherein said receiving further comprising duplicating at least one frame of two packets bordering a gap, said gap detected whenever the sequence number of said two adjacent packets is not consecutive.
- 25 25. A method according to claim 24 wherein said receiving further comprises inserting at least one silence frame between said duplicated frames.
26. A method according to any of claims 24 - 25 wherein said controlling comprises removing at least one frame of the frames duplicated in said step of receiving due to said gap.
27. A method according to any of claims 24 - 26 wherein said controlling comprises duplicating at least one frame of the frames not duplicated by said means for receiving due to said gap.

28. A method according to any of claims 20 - 27 wherein said controlled fullness level is determined in accordance with the variations in size of said playback buffer.
- 5 29. An audio transceiver between an audio device of a personal computer and a packet data network, the audio transceiver comprising:
- a. means for receiving a stream of sequence stamped audio packets from said packet data network; and
  - b. means for duplicating at least one frame of two received packets bordering a gap, said gap detected whenever the sequence number of
- 10 said two adjacent received packets is not consecutive.
30. An audio transceiver according to claim 29 also comprising means for inserting at least one silence frame between said duplicated frames.
31. A method for transmitting data between an audio device of a personal computer and a packet data network, comprising:
- 15 a. receiving a stream of sequence stamped audio packets from said packet data network;
- b. duplicating frames of two received packets bordering a gap, said gap detected whenever the sequence number of said two adjacent received packets is not consecutive.
- 20 32. A method according to claim 31 further comprising inserting at least one silence frame between said duplicated frames.
33. An audio transceiver between an audio device of a personal computer and a packet data network, the audio transceiver comprising:
- a. means for receiving a stream of audio packets from said packet data
- 25 network; and
- b. voice detection means which stores at least a portion of audio packets received before a sound threshold level is reached and which sends after said sound threshold is reached stored audio packets with audio packets received after said threshold value is reached.

34. A method for transmitting and receiving audio data between an audio device of a personal computer and a packet data network, the method comprising the steps of:

- a. receiving a stream of audio packets from said packet data network;
- b. storing at least a portion of audio packets received before a sound threshold level is reached; and
- c. sending after said sound threshold level is reached said stored audio packets with audio packets received after said sound threshold level is reached.



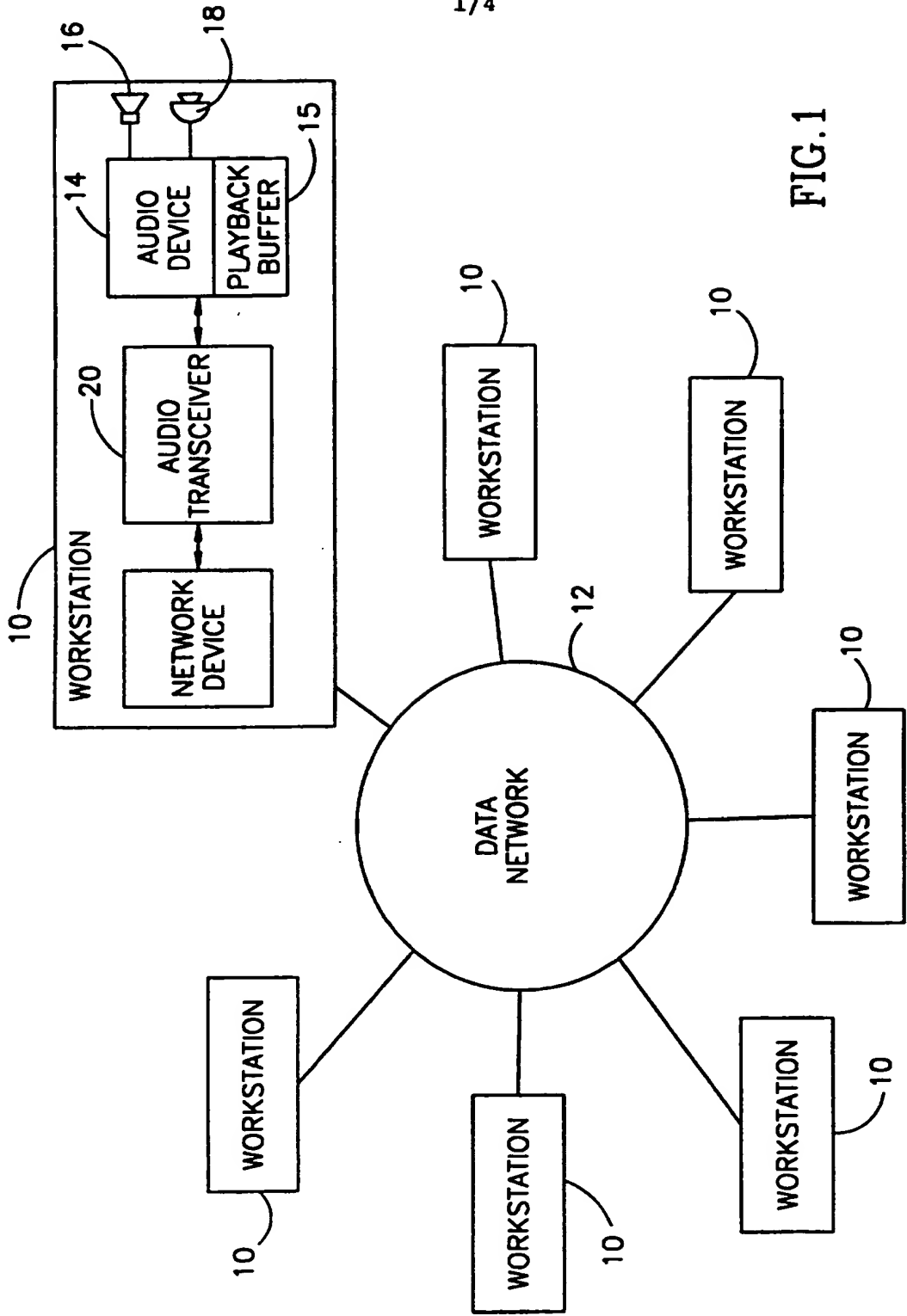


FIG. 1

2/4

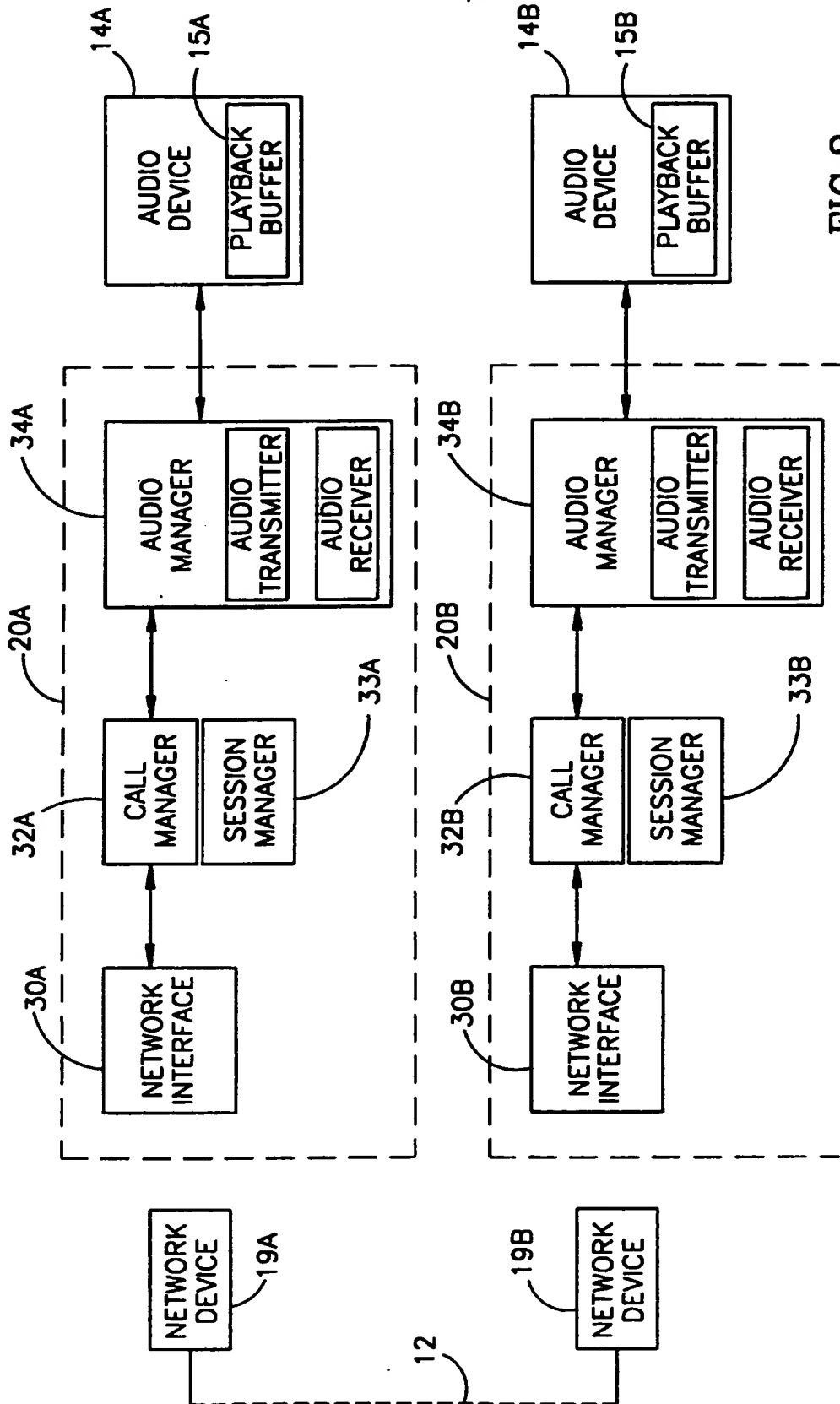
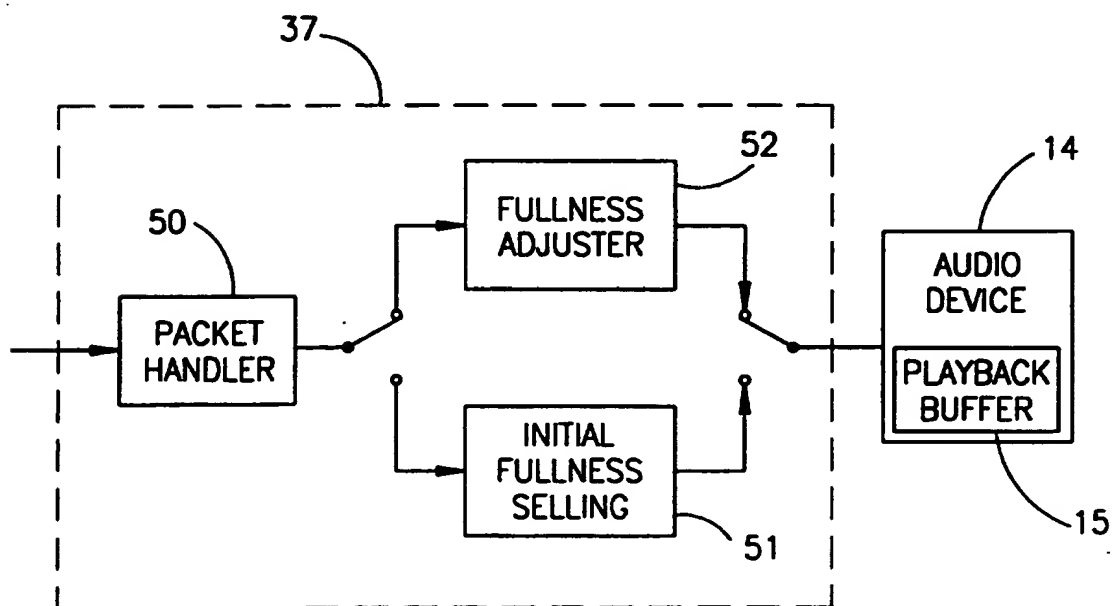
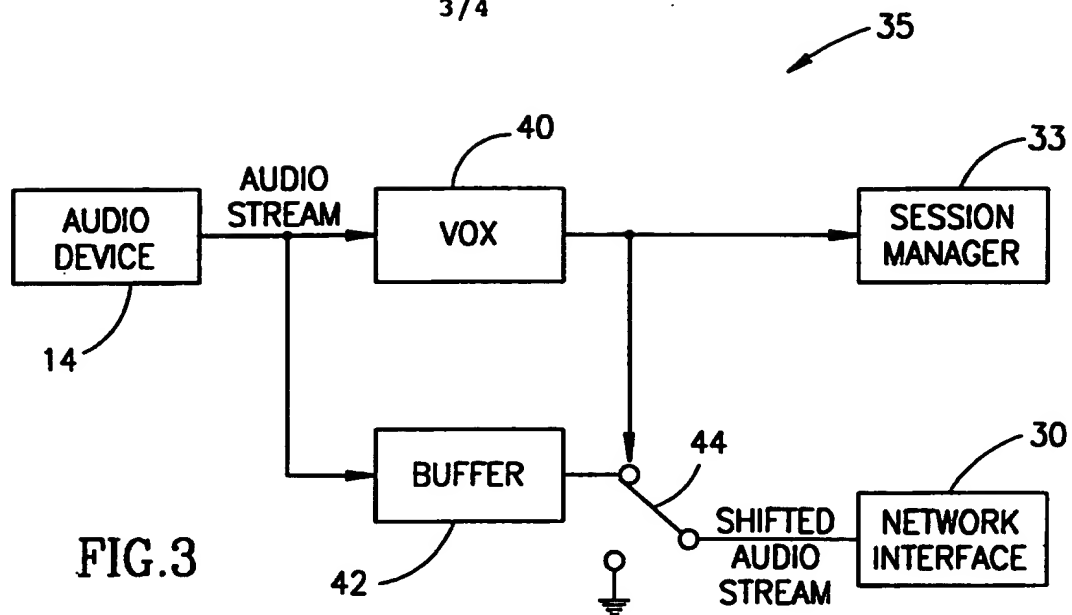


FIG. 2

3/4



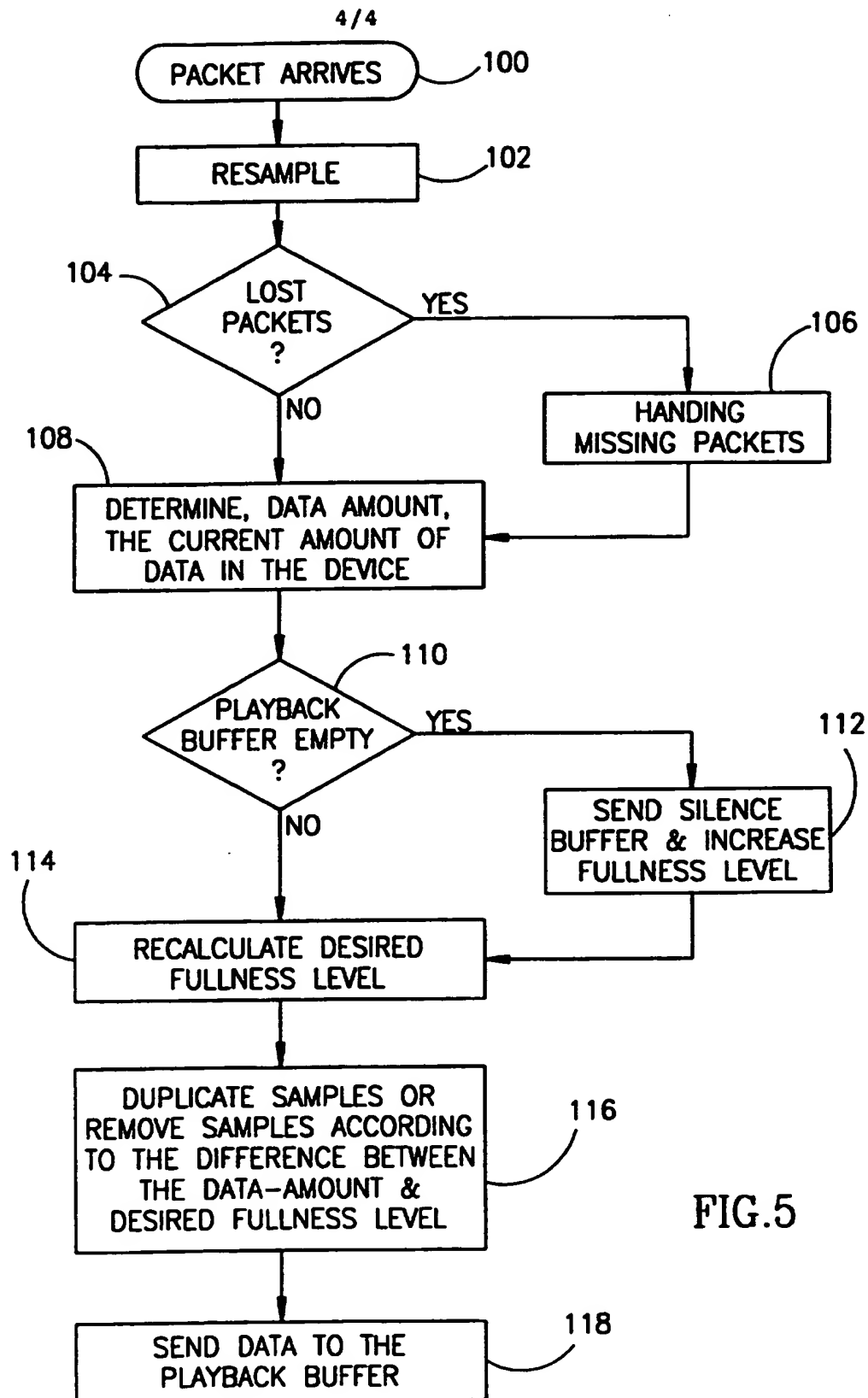


FIG.5

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US95/14123

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) : H04J 3/02, 3/12, 3/16, 3/24; H04Q 11/04

US CL : 370/58.2, 60, 62, 79, 85.1, 94.1, 110.1

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 370/58.2, 60, 62, 79, 85.1, 94.1, 110.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS

search terms: audio, packet, stamp, receiver, playback, buffer

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US, A, 5,323,272 (KLINGLER) 21 June 1994, col. 2, lines 53-63, col. 3, lines 36-55, and col. 3, lines 7-35.	1, 2, 15, 16, 29-34
Y,P	US, A, 5,383,181 (ARAMAKI) 17 January 1995, col. 1, line 47 to col. 2, line 5, col. 26, lines 15-35, col. 5, line 60 to col. 6, line 2, and col. 1, lines 10-19.	1, 2, 15, 16, 29-34
A	US, A, 5,018,136 (GOLLUB) 21 May 1991, see entire document.	1, 2, 15, 16, 29-34
A	US, A, 5,297,139 (OKURA ET AL.) 22 March 1994, see entire document.	1, 2, 15, 16, 29-34
A	US, A, 5,155,760 (JOHNSON ET AL.) 13 October 1992, see entire document.	1, 2, 15, 16, 29-34

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"T" later documents published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be part of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"A" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

26 DECEMBER 1995

Date of mailing of the international search report

01 FEB 1996

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

DOUGLAS OLMS

Telephone No. (703) 305-4703

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/US95/14123

**C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US, A, 4,882,754 (WEAVER ET AL.) 21 November 1989, see entire document.	1, 2, 15, 16, 29-34

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US95/14123

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
  
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
  
3. ☒ Claims Nos.: 4-14, and 18-28  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
  
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
  
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
  
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.